UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

September 26, 2000

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES B-13

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: Donna Kostanich DK
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: Alfredo Navarro AN
Chief, Variance Estimation and Long Form Branch

Subject: Accuracy and Coverage Evaluation Survey: Comparing Accuracy

The attached document describes a general proposal for a report that we will prepare, per your request, following completion of applicable Accuracy and Coverage Evaluation Survey (A.C.E.) operations. The completed report is intended to aid the Executive Steering Committee on A.C.E. Policy (ESCAP) in its recommendation regarding the release of the statistically corrected data or the data without statistical correction as the P.L. 94-171 data. This report, together with other reports, will assess the operations and results of both the initial Census and the A.C.E. Both sets of assessments will be available to the ESCAP to aid the Committee in reaching its recommendation regarding the use of the statistically corrected data.

The attached prototype contains a general description of textual analysis that will assess specific aspects of the applicable operations. This report focuses on the accuracy of corrected population counts as compared to the uncorrected census counts both distributive and numeric. The analysis is limited by assumptions necessary to estimate bias components.

It is important to note that the conduct of the operations may lead us to modify the attached format by including additional information. It is also likely that descriptions and definitions will be enhanced or the data items could undergo revision. Conversely, we may conclude, for a variety of reasons, that some of the information set forth in the attached prototype may not be available. The attached document sets forth our conclusions prior to completion of the A.C.E. about what information would properly inform the ESCAP on this subject, but is subject to modification.

Accuracy and Coverage Evaluation 2000: Comparing Accuracy

prepared by Alfredo Navarro

Introduction

The 2000 Census Data Products Program calls for releasing population counts at the census block level by April 1, 2000. The Census Bureau must decide by this time whether or not these data should be corrected for census net coverage problems. The Census Bureau will be examining many quality indicators for both the Accuracy and Coverage Evaluation (A.C.E.) and the Census to arrive at the decision. This study is one of many that will contribute to the decision process and is unique in that through the use of loss functions, it attempts to pool together individual measures to assess accuracy. This study will compare the accuracy of corrected population counts to the uncorrected census counts primarily focusing on areas roughly the size of congressional districts. Bruce Spencer[1] has suggested using a combination of the 1990 Post Enumeration Survey (PES) evaluation data, the 2000 Census results, and the A.C.E. survey data to assess whether the accuracy is greater for the corrected or uncorrected census. However, a number of assumptions must be made to conduct this analysis, placing limitations on the results.

Background

After the 1990 census, the Census Bureau performed analyses to assess the accuracy of the census and the corrected (or adjusted) census population numbers. Two sets of analyses were conducted over nearly a two-year period to inform on two different adjustment decisions. The first was in preparation for the Secretary of Commerce's July 1991 decision on whether to adjust the 1990 census data. After his decision not to adjust, he requested that the Bureau continue analyzing the PES data to see if technical concerns could be overcome so that the population base for the intercensal population estimates could be adjusted. The overriding question for both decisions was which had more error--the biases and sampling error in the adjusted numbers or the large undercount in the unadjusted census estimates.

The primary assessment of error in the unadjusted census was the PES assessment of undercount. A variety of evaluation studies were carried out to assess possible sources of error in the adjusted numbers. For example, sources included matching errors, fictitious census persons, or missing data. The net effect of the various sources of error is referred to as the "total error" in the adjusted estimate. An analysis was conducted to estimate the mean of the total error (or the overall bias in the adjusted estimate) and the variances and covariances of the adjusted estimates.

---

[1]Spencer, Bruce, "Will Adjustment of Census 2000 Improve Redistricting?", August 1, 2000.

The estimate of total error in the (unadjusted) census numbers was estimated by the difference between the census numbers and the adjusted numbers, after allowance was made for estimated biases in the latter. Explicit allowance was made for the fact that the estimate of total error in the census is subject to variance. The interpretations of the estimates of total error were tempered by knowledge that not all sources of bias in the adjustments were fully reflected in the analysis.

Loss functions are scalar measures of accuracy that summarize the closeness of a set of estimates to the true values. If the true population were known for each geographic area of interest, such as states or counties, then the corrected estimate and the uncorrected census count could be compared to the true population. However, since the true population was unknown, the Census Bureau had to rely on an estimated truth (a target number) to perform loss function analysis. This estimated truth was represented by the adjusted estimates without their associated biases. Direct estimates of bias were available only at the national and evaluation poststrata levels. Therefore, the Bureau had to use models to estimate the level of error for each geographic area of interest.

Given estimates of error for the various areas, the Bureau constructed summary measures of error by applying loss functions such as weighted and unweighted sums of mean squared errors in estimated population figures. Estimates of the difference in expected values of loss functions were subject to sampling error. Hypothesis tests were performed (during the second set of analyses) to see whether the observed difference in expected loss was consistent with a null hypothesis that the unadjusted census was more accurate.

The Bureau almost exclusively conducted analyses using population shares–focusing on distributive accuracy. This was because the effects of adjustment on apportionment were very important to the Secretary's decision as were the effects on Federal and state allocation programs to the Census Bureau Director's recommendation. Although Federal and state allocation programs also have funding levels associated with them, usually the programs allocate a fixed amount of monies among competing governmental units. The analysis was conducted by type of tabulation area, such as: states and places and counties by size categories. These areas compete for resources allocated through these programs.

By August 1992, the Census Bureau Director had to decide whether the population for the intercensal population estimates should be adjusted. Based on the available evidence at this time, the majority of the Census Bureau statisticians concluded that, on average, adjustment improved accuracy at the national level and for states but that results were inconclusive for sub-state areas[2].

However, in late November 1992, results from additional research began to show improvement by adjustment for areas with 100,000 population or more, particularly for counties with 200,000 population or more. These additional results were based on comparisons for large cities and

---

[2]Obenski, S.M. and Fay, R.E., "Analysis of C.A.P.E. Findings on PES Accuracy at Various Geographic Levels", Internal Census Bureau Memorandum, June 6, 2000.

counties and places compared to the balance of states and the balance of counties, respectively. Bureau statisticians began realizing that assessing the impact of adjustment at different geographic levels, the strategy of analyzing shares within a state within a size class, was too restrictive. To discern improvement in adjusted numbers, the undercount levels must differ among the areas of comparison. This was why differences began emerging when large cities, for example, were compared to the balance of the state. These findings are significant because they are based on comparisons that are more relevant to the intended uses of decennial census data, specifically the type of allocation programs previously discussed.

The Census Bureau will not complete many of the A.C.E. evaluations until later in 2001 and 2002. Although A.C.E. variances will be available, complete information on A.C.E. biases will not be. Consequently, the comparison of accuracy for the adjusted (corrected) and unadjusted (uncorrected) Census 2000 population numbers will assume similarity between biases in the A.C.E. and biases in the 1990 PES. If feasible, the PES biases will be modified based on an analysis of differences in 1990 and 2000 data quality indicators, such as differences in unresolved matches. Net undercount estimates will be based on differences between estimates from the A.C.E. and the 2000 census counts, with allowances made for estimates of bias in the A.C.E.

Evaluation Criteria

The following general strategy is planned for estimating the accuracy of the uncorrected and corrected census estimates. The Spencer methodology uses the 1990 PES bias estimates (adjusted for potential changes based on A.C.E. quality indicators) in conjunction with the 2000 Census and A.C.E. results to get less biased estimates of the "truth." We refer to these as targets. Thus, let C denote the uncorrected census, D the corrected census, and B an estimate of bias in D. The target T is defined as $T = D - B$. Let V_B be an estimate of the variance of B, let V_D be an estimate of the variance of D, and suppose the variance of T is $V_T = V_B + V_D$. The mean squared error (MSE) of the census can be estimated by $(C - T)^2 - V_T$ and the MSE of the corrected census can be estimated by $B^2 + V_D - V_B$. The difference in the MSEs, census minus corrected, is estimated as $(C - T)^2 - B^2 - 2V_D$.

Once the target populations are calculated the accuracy of either the census or the corrected census can be established for any tabulation area and for population totals as well as population shares. Note that the targets are calculated using available data from the 1990 PES evaluation studies. The accuracy measures can be calculated for states and 1990 congressional districts in real time, that is, before April 1, 2001. We cannot perform calculations on the congressional districts to be drawn from the 2000 census data, because the districts will not be constructed until after April 1, 2001.

We may be able to conduct comparisons for selected sub-state areas, such as places and counties to provide additional information for the decision. As was true for the 1990 analyses, individual measures of accuracy will be aggregated by area type, such as across congressional districts or states. Additional methodological details and other limitations will be forthcoming. We will

develop a much more detailed implementation and analysis plan, particularly details of the methodology to develop the target populations. Note that the Census Bureau will base its decision on a comprehensive set of data and results, which includes the results of the analysis described in this document.

Additional information on the total error methodology to calculate unbiased measures of accuracy is documented in several papers including *Mulry 1992*[3] and *Mulry and Spencer, 1993*[4]. For a discussion and references to limitations of those measures of accuracy, see the reference in footnote 1. For dissenting views, see for example, Brown, et al[5]

---

[3]Mulry, Mary H., "Loss Function Analysis For The Post Census Review (PCR) Estimates", July 2, 1992.

[4]Mulry, Mary H. and Spencer, Bruce D., "Accuracy of the 1990 Census and Undercount Adjustments", JASA 1993, Vol. 88.

[5]Brown, L.D., et al.,"Statistical Controversies in Census 2000 "; Technical Report 537, Dept. of Statistics, U.C. Berkeley, April 30, 1999.